

IN THE ABSTRACT:

Please replace the original Abstract of the application with the following Abstract:

~~It is, therefore, an object of the present invention to provide a~~ A structure and method for automatically creating a dictionary for clustering text documents, ~~including performing~~ performs a first pass for each of the documents to determine a frequency of each word in each of the documents, ~~creating~~ creates a Hashtable of most frequently occurring words in the documents, ~~performing~~ performs a second pass for each of the documents to determine a frequency of phrases in each of the documents that contain only words in the Hashtable, ~~and adding~~ adds the most frequently occurring phrases to the Hashtable, and ~~outputting~~ outputs the most frequently occurring words and the most frequently occurring phrases as the dictionary. ~~The determination of the frequency of each word can include removing punctuation and case from the documents, removing stop words from the document, replacing words in the documents with synonyms, removing duplicate words from the documents, adding remaining words to the Hashtable, determining the frequency of each word remaining in the Hashtable, and removing words below a frequency level from the Hashtable.~~